AD A099703

DESMATICS, INC.

P. O. Box 618
State College, Pa. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

6

THE PROBLEM OF EXPERIMENTAL DESIGN IN SIMULATION,

by

10

Dennis E. Smith

and

Carl A. Mauro

Presented at:

CORS-TIMS-ORSA Joint National Meeting
Toronto, 3-6 May 1981

14

9 TECHNICAL REPORT NO. 113-4

11 May 1981 12 50

This study was supported by the Office of Naval Research
under Contract No. N00014-79-C-0650 Task No. NR 274-317

15

TABLE OF CONTENTS

		Page
I.	THE PROBLEM OF EXPERIMENTAL DESIGN IN SIMULATION	1
	A. DISCUSSION	2
	B. A FACTOR SCREENING MODEL	3
II.	THREE FACTOR SCREENING APPROACHES	5
	A. EXPERT JUDGMENT	5
	B. GROUP SCREENING	6
	C. RANDOM BALANCE	7
III.	PERFORMANCE MEASURES	9
IV.	SOME PRELIMINARY RESULTS	13
V.	REFERENCES	15

I. THE PROBLEM OF EXPERIMENTAL DESIGN IN SIMULATION

A computer simulation, which is a mathematical model of a system in the form of a computer program, may be viewed as a "black box" in which input factors (independent variables) are combined to produce an output or response (dependent variable). The simulation usually is used to conduct an experimental study of the modelled system. Since simulation runs often are very expensive, the simulation user may wish to concentrate on only the most important factors, that is, those having a strong effect on the output. However, because standard experimental designs found in the statistical literature often require more simulation runs than are available to the simulation user, the identification of these factors by means of statistically designed experiments can pose special design and analysis problems.

In general, therefore, the primary difficulty of experimental design in simulation can be succinctly summarized as too many factors and too few runs. Because of this, it is impossible to investigate thoroughly all factors under consideration. What is required, then, is some means of making the available number of computer runs and the number of factors compatible. Assuming that time and/or budget limitations prohibit additional computer runs, there is a need for conciliatory alternatives that can be feasibly implemented.

A general discussion of this experimental design problem is presented in this report. Three possible two-stage strategies for attacking the problem are considered, and performance measures with which to judge the strategies are described. Each strategy consists of a first stage which uses a nonstandard approach to identify a relatively small factor subset for further consideration. The second stage examines this subset by means of a standard experimental design in an attempt to eliminate any unimportant factors which were unknowingly included in the subset. Because these strategies are designed for "screening" the factors, they are known as factor screening approaches.

A. DISCUSSION

In some cases expert judgment, based on simulation of similar systems or on consideration of the processes being simulated, can be used to select the subset of factors for follow-up experimentation. For example, because of previous experience, the user of a given simulation may be quite certain that specific factors will have little or no effect on the response when compared with the rest of the factors. In this situation, then, these factors could be eliminated from the investigation by keeping them fixed at constant values throughout subsequent experimentation. The remaining factors would comprise the subset to be analyzed in the second stage.

On the other hand, instead of selecting factors according to expectations without expending any computer runs, it may prove of value to invest a portion of the available computer runs in a preliminary first-stage screening experiment. Of necessity, such a preliminary experiment would, as a rule, involve considerably less computer runs than factors, thus giving rise to confounded estimates, that is, estimates which are "mixed together" and impossible to separate by statistical analysis. Confounded estimates provide ambiguous results which may, if interpreted incorrectly, lead to completely erroneous conclusions about which factors are important.

Because unconfounded estimates of the effects of K factors cannot be obtained without a minimum of K + 1 computer runs, the confounding problem may present severe drawbacks to the usefulness of any preliminary screening experiment. Nonetheless, they may prove of enough value to be used instead of (or, possibly, in conjunction with) expert judgment.

B. A FACTOR SCREENING MODEL

In screening, a small number of factor levels is generally employed; usually two are sufficient. Suppose, then, that a simulation consists of K factors, each of which is at two levels, arbitrarily designated "high" and "low."

The actual functional or statistical relationship between the simulation response and the factors of a simulation model will, of course, vary from model to model. However, in devising factor screening strategies for use in computer simulation experiments, it is desirable to define a common statistical model to serve as a basis in which to compare and to assess any screening strategies that might be proposed. To that end, the following paragraphs summarize a reasonable and generally adequate screening model that will be assumed to underlie the simulation responses.

Define

$$x_{ij} = \begin{cases} +1, & \text{if factor } j \text{ is at its "high" level for the } i^{\underline{th}} \text{ computer run} \\ -1, & \text{if factor } j \text{ is at its "low" level for the } i^{\underline{th}} \text{ computer run} \end{cases}$$

and let y_i denote the simulation response for the $i^{\underline{th}}$ computer run. The factor screening model assumes that

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} + \varepsilon_i,$$

where β_j is the (linear) effect of factor j and the error terms, ε_i, are independent and normally distributed random variables having a zero mean and variance σ^2. In essence, this model may be regarded as a first-order Taylor series approximation to the actual relationship between the y_i's and the x_{ij}'s.

In terms of the model, factor j will be termed active if and only if $\beta_j \neq 0$, and inactive if and only if $\beta_j = 0$. Furthermore, under the adopted parameterization, β_j can be interpreted as the average difference between the true simulation responses of the high and of the low levels of the j^{th} factor. Hence $\beta_j > 0$ only if the factor level producing the larger true response is labeled as the high (+1) level. It is assumed that only a relatively small number, k, of the K factors are active.

Under this nomenclature, the basic aim of any screening procedure is to efficiently and effectively classify, as active or as inactive, the K factors under investigation.

II. THREE FACTOR SCREENING APPROAHCES

This paper considers three possible factor screening approaches. Each approach is a two-stage strategy which combines a nonstandard first stage procedure with a second stage that employs a standard experimental design known as a Plackett-Burman design. [See Plackett and Burman (1946).] This design is a two-level orthogonal design for studying up to 4m-1 factors in 4m runs. Because of the orthogonality, there is no confounding (i.e., mixing together) of factor effects in the second stage.

A. EXPERT JUDGMENT

The first approach assumes that the analyst (i.e., the simulation user) feels he or she can do a good job of deciding which factors are active and which are inactive. Thus, the analyst will, using expert judgment, select those factors to be carried over into the second stage. Assume, for sake of analysis simplicity, that

(1) P(Analyst identifies a factor as active|the factor is active) = r_1

and (2) P(Analyst identifies a factor as inactive|the factor is inactive) = r_2.

Of course, if $r_1 = r_2 = 1.0$, the analyst's judgment is perfect. However, as the probabilities r_1 and r_2 decrease from 1.0, the effectiveness of this method also decreases. Although the second stage Plackett-Burman design applied to factors selected in the first stage helps guard against the misclassification of inactive factors, any active factor not selected by the analyst in the first stage will never be classified correctly.

B. GROUP SCREENING

Group screening has been discussed in a number of papers [e.g., Watson (1961), Li (1962), Mauro and Smith (1980)]. In group screening, "group-factors" are created by partitioning the individual factors into a number of groups. The two-stage group screening procedure considered here relies on a Plackett-Burman design to test for significant group-factor effects in the initial stage. However, this design is used in a nonstandard manner since all factors in a given group appear at the same level during a simulation run.

For example, suppose that the m factors $x_1,\ldots,x_m$ form one group-factor. Then, whenever this particular group-factor appears at its high (+1) level in the Plackett-Burman design, all component factors $x_1,x_2,\ldots,x_m$ would be at their high levels. Thus, the effects of $x_1,x_2,\ldots,x_m$ are completely confounded so that if the group factor is found to have an effect, it cannot be determined which of the factors $x_1,x_2,\ldots,x_m$ or how many of have an effect. The second stage Plackett-Burman follow-up, therefore, helps to resolve this question by examining all individual factors comprising the group factors judged significant in the first stage.

The first stage experiment requires N runs, where N is the smallest integer which is a multiple of four and also greater than the number of group factors. Furthermore, unlike the expert judgment approach, group screening examines all of the original K factors experimentally; none are excluded from experimentation in the first stage. However, the possibility of cancellation of effects within a group factor exists. That is, individual factors could possibly have offsetting positive and negative effects. In such a case, these factors would not be brought over into the second stage and would therefore be misclassified as inactive.

Because of this possibility, the definition of high and low factor levels should be made so that all factor effects are anticipated to have the same direction, e.g., to all be positive. If all the effects have the same direction, cancellation is impossible. Mauro and Smith (1980) have examined, in the case $\sigma = 0$, the performance of group screening when some effect directions are incorrectly assumed.

C. RANDOM BALANCE

In the random balance approach, all K original factors are included in a first stage experiment of N runs. Because of the constraints on the number of runs, $N<K$. Subject to this restriction, the value of N can be whatever the analyst chooses, except that it should be an even number.

In the initial experiment, each factor appears at its high level N/2 times and at its low level N/2 times during the N runs, with the order of high and low levels selected at random. Although this guarantees that the factor effects are unconfounded with the overall mean effect, they are confounded with each other. Furthermore, the confounding is random. In addition, no standard analysis techniques for random balance data exist, although a number have been suggested.

However, proponents of random balance [e.g., Satterthwaite (1959) and Budne (1959)] have emphasized that, in general, the degree of confounding is relatively small and analysis poses no great problem. Nonetheless, random balance has received a very bad name in the statistical community, mainly because of the random confounding of factor effects. Although the objections are based on good statistical reasoning, no empirical evidence is available to support either proponents or opponents of random balance. Mauro and Smith (1981) are currently investigating the performance of random balance when a standard one-factor analysis of variance F-test is used as the method of analysis. The second stage Plackett-Burman design includes all factors judged significant in the random balance experiment.

III. PERFORMANCE MEASURES

In attempting to identify the important factors for detailed investigation, there are the two conflicting requirements of factor misclassification and expenditure of runs. Before different factor screening approaches may be compared, these requirements must be quantified. In assessing performance, Smith and Mauro (1980) considered the values of expected loss and expected relative testing cost.

In order to measure the severity of classification error, consider the class of loss functions given by

$$L = \sum_{1}^{K} w_j \delta_j \Big/ \sum_{1}^{K} w_j$$

where

$$\delta_j = \begin{cases} 0, & \text{if the } j^{\underline{th}} \text{ factor is correctly identified} \\ 1, & \text{if the } j^{\underline{th}} \text{ factor is incorrectly identified,} \end{cases}$$

and w_j denotes the loss incurred ($w_j \geq 0$) if the $j^{\underline{th}}$ factor is misclassified. Note that L is a function of $\beta_1, \ldots, \beta_K$ and lies in the interval [0,1].

For the particular case in which

$$|\beta_j| = \begin{cases} \Delta, & \text{if factor } j \text{ is active} \\ 0, & \text{if factor } j \text{ is inactive,} \end{cases}$$

it is reasonable to let

$$w_j = \begin{cases} 1/2k, & \text{if factor } j \text{ is active} \\ 1/2(K-k), & \text{if factor } j \text{ is inactive,} \end{cases}$$

since this apportions one-half of the overall maximum loss to the active factors and the other half to the inactive factors. Hence, in this case the loss L reduces to

$$L = [(K-k)(k-A) + k(K-k-I)]/2k(K-k)$$

where A denotes the number of active factors correctly identified and I denotes the number of inactive factors correctly identified.

The second performance measure discussed by Smith and Mauro (1980) takes into account the total number of runs, R, that a factor screening approach requires. The testing cost may be defined relative to the number of runs required for a Plackett-Burman design applied to all K original factors. Thus, the relative testing cost Q is given by

$$Q = \phi(R)/\phi(K^*)$$

where $\phi(M)$ represents the expense of conducting M runs, and K^* denotes the number of runs required by a Plackett-Burman design for K factors. If $\phi(M)$ is assumed proportional to M, then

$$Q = R/K^*.$$

It should be noted that in most screening strategies both L and Q are random variables. Thus, in assessing the performance of a factor screening approach, it is reasonable to examine their expected values.

Both expected loss and expected relative testing cost must be jointly considered in evaluating the overall performance of a factor screening strategy. In some sense the problem is akin to the testing of a statistical hypothesis in which the probabilities of Type I error (rejecting a true null hypothesis) and Type II error (accepting a false null hypothesis) are both desired small, but are inversely related.

The simulation user may wish to specify joint values of expected loss and expected relative testing cost that are acceptable. For example, the user may place an upper limit on expected loss and then, subject to this constraint, select the screening approach having the minimum relative testing cost.

Only if one screening strategy has both a smaller expected loss, E(L), and expected relative testing cost, E(Q), than another strategy can the first be said to be definitely better than the second. Otherwise, the decision depends upon the analyst's trade-offs. For example, by looking at Figure 1, it is clear that all analysts would select strategy A over either strategy B or D. However, one analyst might prefer A over E because of the smaller E(Q) while another might prefer E over A because of the smaller E(L).

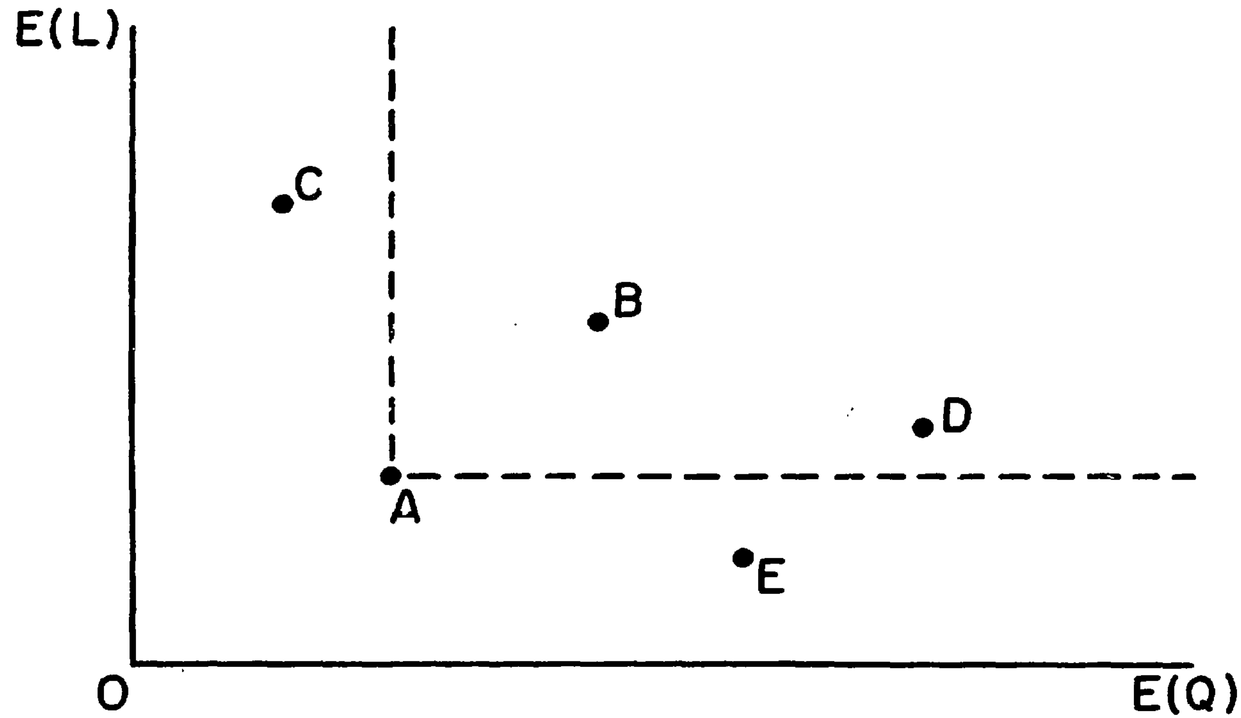

Figure 1: Required Trade-offs for selecting a Screening Strategy

IV. SOME PRELIMINARY RESULTS

Ongoing research by Desmatics, Inc. is examining the performance measures E(L) and E(Q) for the situation where all active factors are such that $|\beta_j| = \Delta$, $\phi(M)$ is proportional to M, and the incurred loss w_j is as defined in the previous section. Within this framework, the following cases are being considered:

$K = 60, 120, 240$

$k = p^*K \qquad (p^* = 2/60, 3/60, 5/60, 8/60)$

$\sigma = r\Delta \qquad (r=0, r>0)$

Research to date has considered only the deterministic case (i.e., r=0). Future research will address the case where random error is present.

Figure 2 exhibits results for the specific case K = 120, k = 10, and $\sigma = 0$. In the deterministic situation, E(A), the expected number of active factors identified, is equal to k[1-2E(L)] for the three approaches considered in this report. Thus, both E(L) and E(A) are presented in the figure.

As will be noted, there are a number of points corresponding to each of the three strategies. For the expert judgment strategy, performance depends on the values of the probabilities r_1 and r_2. The figure gives results for various values of $r_1 = r_2$. For group screening, performance depends on g, group size, and on i, the number of misspecified factor effect directions. The figure provides results for g = 3, 5, and 8 and i = 0, 1, 2, 3, 4, 5. For random balance, performance depends on c, where c = N/K and on α, the significance level for the F-test used in analyzing the first-stage data. The results in Figure 2 correspond to various values of c and α in the ranges $.2 < c < .8$ and $.10 < \alpha < .50$.

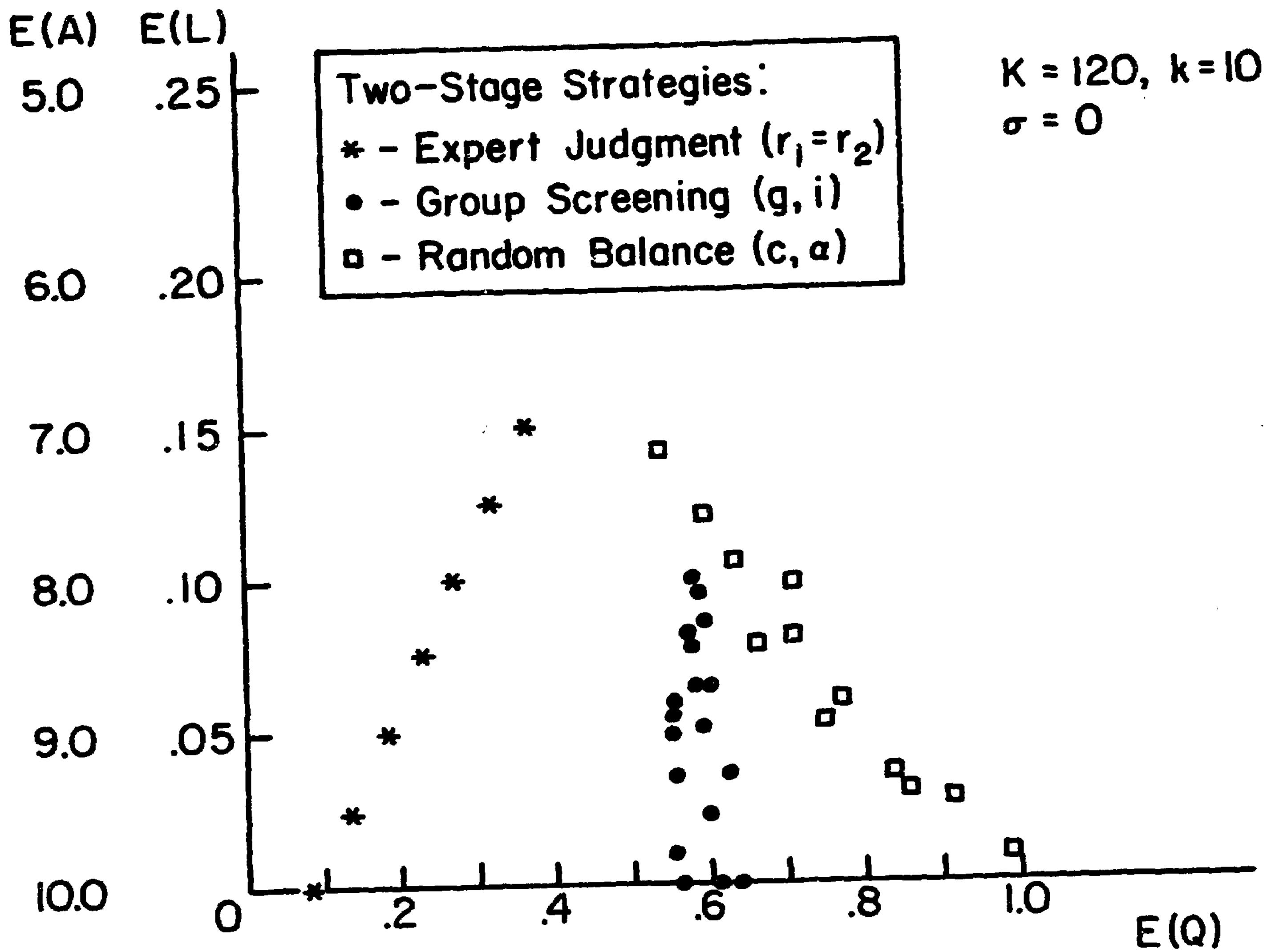

Figure 2: Some Preliminary Results

V. REFERENCES

Budne, T.A. (1959). "The Application of Random Balance Designs," Technometrics, Vol. 1, p. 139.

Li, C.H. (1962). "A Sequential Method for Screening Experimental Variables," Journal of the American Statistical Association, Vol. 57, p. 455.

Mauro, C.A., and Smith, D.E. (1980). "An Examination of the Performance of Two-Stage Group Screening for Use in Computer Simulation," Technical Report No. 113-3, Desmatics, Inc.

Mauro, C.A., and Smith, D.E. (1981). "An Examination of the Performance of Random Balance Screening for Use in Computer Simulation," Technical Report No. 113-5 (in preparation), Desmatics, Inc.

Plackett, R.L., and Burman, J.P. (1946). "The Design of Optimum Multifactor Experiments," Biometrika, Vol. 33, p. 305.

Satterthwaite, F.E. (1959). "Random Balance Experimentation," Technometrics, Vol. 1, p. 111.

Smith, D.E., and Mauro, C.A. (1980). "Factor Screening in Computer Simulation: Considerations in Performance Evaluation," Technical Report No. 113-2, Desmatics, Inc.

Watson, G.S. (1961). "A Study of the Group Screening Method," Technometrics, Vol. 3, p. 371.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 113-4	2. GOVT ACCESSION NO. AD-A099 703	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE PROBLEM OF EXPERIMENTAL DESIGN IN SIMULATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dennis E. Smith and Carl A. Mauro		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0650
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P.O. Box 618 State College, PA 16801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 274-317
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE May 1981
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Experimental Design
Computer Simulation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In many cases, a simulation study may be viewed as an experimental situation in which a number of factors (independent variables) are to be investigated. However, standard experimental design techniques often require more simulation runs than are available to the simulation user. In general, the primary problem of experimental design in simulation can be succinctly summarized as too many factors and too few runs. A discussion cont

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(continuation of 20.)

cont of this problem is presented in this report. Three possible two-stage strategies for attacking the problem are considered, and performance measures with which to judge the strategies are described.